

PERBANDINGAN METODE KLASIFIKASI ARTIFICIAL NEURAL NETWORK BACKPROPAGATION DAN REGRESI LOGISTIK

(Studi Kasus : Bank Internasional Indonesia)

Siti Hadijah Hasanah

Program Studi Manajemen, Fakultas Ekonomi - UNPAM

dosen02011@unpam.ac.id

ABSTRACT

Unsecured loan is one of the credit product from given bank to customers in the form of a loan facility without a guarantee. Because there is no collateral for the loan, so the bank must carefully examine the credit rating of the customer in order to avoid the risk of losses in the future. Application submission of unsecured loan by the customer to the bank will be assessed based on classification techniques. Classification techniques unsecured loan on using a statistical approach is logistic regression and neural network (ANN). Logistic regression is one of the parametric method which not required a the assumptions must be met when analyzing data using linear regression. ANN method is processing information that is inspired by biological nervous system (Haykin, 1999). Logistic regression has the ability to determine the explanatory variables that influence the response variable result. Logistic regression with explanatory variables that influence is gender, number of installments 12 month, number of installments 24 month, and salary. So the bank can make the explanatory variables such as consideration for determining the customer's decision results KTA. Based on the accuracy the classification of confusion matrix, the accuracy value and AUC in training data and testing data. The best method on unsecured loan customer data is ANN Backpropagation followed logistic regression.

Keywords: *Unsecured Loan, Logistic Regression, Artificial Neural Network Backpropagation.*

ABSTRAK

Kredit tanpa agunan (KTA) adalah salah satu produk kredit yang diberikan bank kepada nasabah kredit dalam bentuk fasilitas pinjaman tanpa ada suatu jaminan. Karena tidak ada jaminan atas pinjaman tersebut maka bank harus berhati-hati memeriksa calon nasabah kredit agar tidak terjadi resiko kerugian di kemudian hari. Pengajuan aplikasi KTA oleh nasabah kepada pihak bank akan dilakukan penilaian berdasarkan teknik klasifikasi. Teknik klasifikasi pada KTA ini menggunakan metode pendekatan statistik yaitu regresi logistik dan ANN. Regresi logistik merupakan salah satu metode parametrik yang tidak disyaratkan asumsi-asumsi sebagaimana yang harus dipenuhi apabila melakukan analisis data dengan menggunakan regresi linear. Metode ANN adalah pemrosesan informasi yang terinspirasi oleh sistem syaraf biologi (Haykin,1999). Metode regresi logistik memiliki kemampuan untuk menentukan peubah penjelas yang

berpengaruh terhadap peubah respon hasil keputusan. Regresi logistik dengan peubah penjelas berpengaruh yaitu jenis kelamin, jumlah cicilan 12 bulan, jumlah cicilan 24 bulan, dan standar gaji. Jadi pihak bank dapat menjadikan peubah penjelas tersebut sebagai pertimbangan untuk menentukan hasil keputusan nasabah KTA. Berdasarkan nilai ketepatan klasifikasi *confusion matrix*, nilai akurasi, dan AUC pada data *training* dan data *testing* metode yang terbaik pada data nasabah KTA yaitu ANN *Backpropagation* diikuti oleh regresi logistik.

Kata Kunci : Kredit Tanpa Agunan, *Artificial Neural Network* (ANN) *Backpropagation*, Regresi Logistik.

1. PENDAHULUAN

1.1. Latar Belakang Masalah

Teknik klasifikasi merupakan salah satu metode statistik dalam pengelompokan suatu data yang disusun secara sistematis. Teknik klasifikasi yang telah dikembangkan sampai saat ini dengan menggunakan metode pendekatan statistik seperti analisis diskriminan, regresi logistik, *multivariate adaptive regression splines* (MARS), *classification and regression tree* (CART), *k-nearest neighbours* (KNN), *support vector machines* (SVM), dan *artificial neural network* (ANN) (Mirtalaei *et al.* 2012). Regresi logistik merupakan salah satu metode parametrik yang digunakan untuk menduga hubungan antara peubah respon kategori dengan satu atau lebih peubah penjelas yang kontinu maupun kategori. Regresi logistik tidak disyaratkan asumsi-asumsi sebagaimana yang harus dipenuhi apabila melakukan analisis data dengan menggunakan regresi linear. Metode lain yang dapat digunakan dan merupakan suatu teknik klasifikasi yang fleksibel dan inovatif seperti metode *artificial neural network* (ANN).

Metode ANN adalah pemrosesan informasi yang terinspirasi oleh sistem syaraf biologi (Haykin, 1999). Kelebihan ANN adalah dapat memodelkan hubungan yang nonlinear, memiliki kemampuan untuk belajar (adaptif) sehingga dapat mempresentasikan pengetahuan secara flexibel terhadap adanya kesalahan. ANN *Backpropagation* merupakan algoritma pembelajaran terawasi yang terdiri dari lapisan *input*, lapisan tersembunyi, dan lapisan *output* dengan mengubah bobot-bobot yang terhubung pada masing-masing lapisan (Shi *et al.* 2012). Algoritma *Backpropagation* menggunakan sisaan keluaran untuk mengubah nilai bobot pada tahap maju (*forward*) dan tahap mundur (*backward*).

Teknik klasifikasi dapat diterapkan pada pengklasifikasian kemiskinan dan masalah perbankan. Pertumbuhan perbankan yang cukup pesat di Indonesia menimbulkan adanya persaingan antar bank untuk saling memberikan pelayanan yang terbaik kepada para nasabah. Salah satu pelayanan perbankan yang cukup diminati oleh nasabah adalah pemberian kredit. Kredit tanpa agunan (KTA) adalah salah satu produk kredit yang diberikan bank kepada nasabah kredit dalam bentuk fasilitas pinjaman tanpa ada suatu jaminan. Karena tidak ada jaminan atas pinjaman tersebut maka bank harus berhati-hati memeriksa *rating* kredit para nasabah agar tidak terjadi resiko kerugian di

kemudian hari. Penelitian ini untuk membandingkan metode klasifikasi antara regresi logistik dan ANN *Backpropagation*, berdasarkan ketepatan klasifikasi dengan kurva *receiver operating characteristic* (ROC) dan *confusion matrix*.

1.2. Rumusan Masalah

Bagaimana melakukan perbandingan pada metode regresi logistik dan *Artificial Neural Network Backpropagation* dalam pengklasifikasian data.

1.3. Tujuan Penelitian

Tujuan penelitian adalah melakukan perbandingan pada metode regresi logistik dan *Artificial Neural Network Backpropagation* dalam pengklasifikasian data.

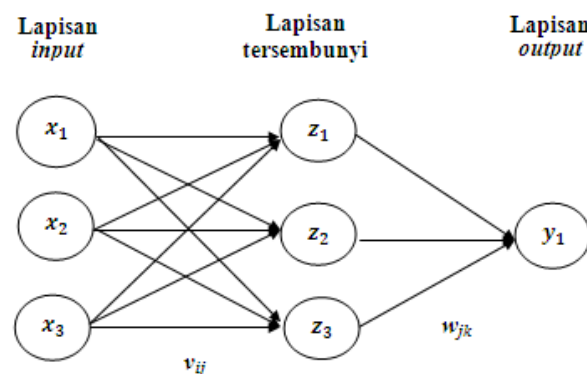
2. METODOLOGI PENELITIAN

2.1. Artificial Neural Network (ANN)

Metode ANN adalah pemrosesan informasi yang terinspirasi oleh sistem syaraf biologi (Haykin, 1999). ANN mampu mengenali kegiatan dengan berbasis data pada masa lalu. Data masa lalu akan dipelajari oleh ANN sehingga mempunyai kemampuan untuk memberi keputusan terhadap data yang belum pernah dipelajari.

Lapisan-lapisan penyusun ANN dapat dibagi menjadi tiga yaitu :

1. Lapisan input
Neuron di dalam lapisan *input* disebut unit-unit *input*. Unit-unit *input* menerima *input* dari dunia luar.
2. Lapisan tersembunyi
Neuron di dalam lapisan tersembunyi disebut unit-unit tersembunyi. *Output* dari lapisan ini tidak secara langsung dapat diamati
3. Lapisan output
Neuron pada lapisan *output* disebut unit-unit *output*. *Output* dari lapisan ini merupakan *output* ANN terhadap suatu permasalahan.



Gambar 1 Arsitektur *Artificial Neural Network* (Sumber : Reyson, 2012)

Gambar 1 merupakan salah satu contoh arsitektur ANN yang terdiri dari satu lapisan *input*, satu lapisan tersembunyi, satu lapisan *output*, v_{ij} adalah bobot antara

lapisan *input* dengan lapisan tersembunyi ($i = 1, 2, 3$) dan ($j = 1, 2, 3$), dan w_{jk} adalah bobot antara lapisan tersembunyi dengan lapisan *output* ($j = 1, 2, 3$) dan ($k = 1$).

Proses *training* pada ANN melatih jaringan untuk mendapatkan keseimbangan antara kemampuan jaringan dalam mengenali pola serta kemampuan jaringan untuk memberikan respon yang benar terhadap pola *input* yang serupa dengan pola yang dipakai selama pelatihan. Proses pembelajaran dalam ANN terbagi dua yaitu :

1. *Supervised learning* merupakan proses pembelajaran yang terawasi dimana *output* yang diharapkan telah diketahui sebelumnya. Biasanya pembelajaran ini dilakukan dengan menggunakan data yang telah ada.
2. *Unsupervised learning* merupakan pembelajaran yang tidak terawasi dimana tidak memerlukan target *output*.

2.1.1. *Backpropagation*

Backpropagation merupakan algoritma pembelajaran terawasi yang terdiri dari lapisan *input*, lapisan tersembunyi, dan lapisan *output* dengan mengubah bobot-bobot yang terhubung pada masing-masing lapisan (Shi *et al.* 2012).

Algoritma ANN *Backpropagation* dapat dibagi dalam 2 bagian :

1. Algoritma pelatihan

Terdiri dari 3 tahap yaitu :

- a. Tahap maju

Pola *input* dihitung maju mulai dari *lapisan input* hingga *lapisan output* menggunakan fungsi aktivasi yang ditentukan.

- b. Tahap mundur

Selisih antara *output* jaringan dengan target yang diinginkan merupakan kesalahan yang terjadi. Kesalahan yang terjadi itu dilakukan proses mundur, dimulai dari garis yang berhubungan langsung dengan unit-unit di *lapisan output*.

- c. Tahap pengaturan bobot

Modifikasi bobot untuk menurunkan kesalahan yang terjadi. Ketiga tahap tersebut diulang terus-menerus hingga kondisi penghentian dipenuhi.

2. Algoritma aplikasi

Tahap maju yaitu pola *input* dihitung maju mulai dari *lapisan input* hingga *lapisan output* menggunakan fungsi aktivasi yang ditentukan.

Model ANN *Backpropagation* adalah sebagai berikut :

$$y_k = f_k \left(\sum_{j=1}^p w_{jk} f_j \left(v_{0j} + \sum_{i=1}^n x_i v_{ij} \right) + w_{0k} \right) \quad (1)$$

w_{0k} = bobot bias pada unit *output* y_k

v_{0j} = bobot bias pada unit tersembunyi z_j

v_{ij} = bobot garis dari unit x_i ke unit tersembunyi z_j

w_{jk} = bobot garis dari z_j ke unit *output* y_k

f_k = fungsi aktivasi pada unit tersembunyi ke *output*

f_j = fungsi aktivasi pada unit *input* ke tersembunyi

x_i = unit *input* ke- i

z_j = unit tersembunyi ke- j

y_k = unit *output* ke- k

i = 1, ..., n

j = 1, ..., p

k = 1, ..., l

n = banyaknya unit *input*

p = banyaknya unit tersembunyi

l = banyaknya unit *output*

2.1.2. Fungsi Aktivasi

Fungsi aktivasi merupakan suatu fungsi yang digunakan untuk membawa nilai *input* menuju *output* yang diinginkan. Fungsi ini harus memenuhi beberapa syarat yaitu kontinu, terdiferensial dengan mudah, dan fungsi yang tidak turun. Menurut (Mirtalaei *et al.* 2012) fungsi aktivasi yang digunakan dalam ANN *Backpropagation* yaitu fungsi sigmoid biner, fungsi tangen sigmoid, dan fungsi linear.

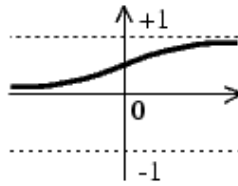
1. Fungsi sigmoid biner

$$y = f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

dengan turunan pertama :

$$f'(x) = f(x)(1 - f(x)) \quad (3)$$

Fungsi sigmoid biner ini diilustrasikan pada Gambar 2 :



Gambar 2 Fungsi sigmoid biner

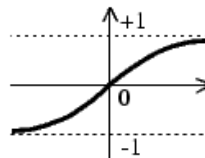
2. Fungsi tangen sigmoid

$$y = f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (4)$$

dengan turunan:

$$f'(x) = (1 + f(x))(1 - f(x)) \quad (5)$$

Fungsi tangen sigmoid digambarkan pada Gambar 3.



Gambar 3 Fungsi tangen sigmoid

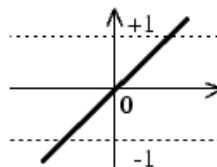
3. Fungsi linear

$$y = f(x) = x \quad (6)$$

dengan turunan :

$$f'(x) = 1 \quad (7)$$

Fungsi linear digambarkan pada Gambar 4.



Gambar 4 Fungsi linear

Algoritma pelatihan *Backpropagation* adalah sebagai berikut:

Tahap maju

1. Tentukan bobot lapisan dengan bilangan acak kecil, -0.5 s/d 0.5.
2. Tiap unit *input* (x_i) mengirimkan sinyal ke unit tersembunyi (z_j). Hitung semua nilai unit tersembunyi (z_j) dengan fungsi aktivasi sigmoid biner.

$$z_j = f_j(v_{0j} + \sum_{i=1}^n x_i v_{ij})$$

z_j = hasil keluaran unit tersembunyi

3. Tiap unit tersembunyi (z_j) mengirimkan sinyal ke unit *output* (y_k). Hitung semua nilai unit *output* (y_k) dengan fungsi aktivasi sigmoid biner.

$$y_k = f_k(w_{0k} + \sum_{j=1}^p z_j w_{jk})$$

y_k = hasil keluaran unit *output*

Tahap mundur

4. Setiap unit *output* y_k menerima pola target t_k selanjutnya, hitung informasi kesalahan lapisan *output* δ_k .

$$\delta_k = (t_k - y_k) f'(y_k)$$

$$= (t_k - y_k) y_k (1 - y_k)$$

δ_k = informasi kesalahan lapisan *output*

t_k = kategori aktual

5. Hitung besar koreksi bias dan bobot Δw_{0k} dan Δw_{jk} antara lapisan tersembunyi dengan lapisan *output*.

$$\Delta w_{0k} = \alpha \delta_k$$

$$\Delta w_{jk} = \alpha \delta_k z_j$$

α = *learning rate*

6. Setiap unit di lapisan tersembunyi dilakukan perhitungan informasi kesalahan lapisan tersembunyi (δ_j).

$$\delta_j = \sum_{k=1}^l \delta_k w_{jk} f'(z_j)$$

$$= \sum_{k=1}^l \delta_k w_{jk} z_j (1 - z_j)$$

δ_j = informasi kesalahan lapisan tersembunyi

7. Hitung besar koreksi bias Δv_{0j} dan bobot Δv_{ij} antara lapisan *input* dan lapisan tersembunyi :

$$\Delta v_{0j} = \alpha \delta_j$$

$$\Delta v_{ij} = \alpha \delta_j x_i$$

Tahap pengaturan bobot

8. Hitung semua perubahan bias dan bobot :

Perubahan bias dan bobot unit *output* :

$$w_{jk}(\text{baru}) = w_{jk}(\text{lama}) + \Delta w_{jk}$$

Perubahan bias bobot unit tersembunyi:

$$v_{ij}(\text{baru}) = v_{ij}(\text{lama}) + \Delta v_{ij}$$

9. Uji syarat berhenti :

Jika besar kuadrat tengah galat (KTG) $\sum_{k=1}^l (t_k - y_k)^2$ lebih kecil dari toleransi yang telah ditentukan atau jumlah *epoch* pelatihan sudah mencapai maksimum maka selesai, jika tidak maka kembali ke langkah 1.

2.1.3. Gradient Descent

Jaringan pada proses pelatihan dilakukan dalam rangka pengaturan bobot sehingga pada akhir pelatihan akan diperoleh bobot-bobot yang baik. Selama proses pelatihan, bobot-bobot diatur secara iteratif untuk meminimumkan fungsi kinerja jaringan. Fungsi kinerja jaringan yang digunakan pada *Backpropagation* adalah kuadrat tengah galat (KTG), fungsi ini akan mengambil rata-rata kuadrat sisaan yang terjadi antara *output* jaringan dan target.

Algoritma proses pelatihan menggunakan *gradient* dari fungsi kinerja yang digunakan untuk menentukan bagaimana mengatur bobot-bobot dalam rangka meminimumkan kinerja yang disebut dengan *gradient descent*. *Gradient* ini biasanya menggunakan nilai *learning rate* (α) yang cukup kecil agar langkah pengulangan yang dilakukan tidak terlalu besar, sehingga didapatkan perkiraan *gradient descent* yang mendekati nilai sebenarnya. Persamaan *learning rate* (α) pada algoritma *Backpropagation* sebagai berikut :

$$\Delta w_{0k} = \alpha \delta_k \quad (8)$$

$$\Delta w_{jk} = \alpha \delta_k z_j \quad (9)$$

dan

$$\Delta v_{0j} = \alpha \delta_j \quad (10)$$

$$\Delta v_{ij} = \alpha \delta_j x_i \quad (11)$$

2.2. Regresi Logistik

Regresi logistik adalah salah satu model untuk menduga hubungan antara peubah respon kategori dengan satu atau lebih peubah penjelas yang kontinu maupun kategori. Misalkan n peubah bebas berskala interval $\mathbf{x}' = (x_1, x_2, \dots, x_n)$ yang merupakan vektor peubah bebas, dan peluang bersyarat $P(Y = 1 | \mathbf{x}) = \pi(\mathbf{x})$, maka bentuk model regresi logistik berganda (Hosmer dan Lemeshow 2000) adalah sebagai berikut :

$$\pi(\mathbf{x}) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)} \quad (12)$$

Nilai peluang respon $\pi(\mathbf{x})$ dengan $0 \leq \pi(\mathbf{x}) \leq 1$, maka model logistik menggambarkan suatu peluang dengan mentransformasikan $\pi(\mathbf{x})$ dengan transformasi logit sebagai berikut :

$$g(\mathbf{x}) = \ln \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] \quad (13)$$

Diperoleh bentuk logit sebagai berikut :

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (14)$$

2.2.1. Pendugaan Parameter Regresi Logistik

Misalkan suatu contoh yang terdiri dari N pengamatan berpasangan (\mathbf{x}_i, y_i) , yang bebas satu sama lainnya, pendugaan model regresi logistik berganda diharuskan memperoleh dugaan vektor $\boldsymbol{\beta}' = (\beta_0, \beta_1, \dots, \beta_n)$. Metode pendugaan yang dapat digunakan pada peubah bebas ganda salah satunya adalah metode kemungkinan maksimum. Fungsi kemungkinan maksimum dapat dituliskan sebagai berikut :

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i} \quad (15)$$

$$\ln[l(\boldsymbol{\beta})] = \sum_{i=1}^n \{y_i \ln [\pi(\mathbf{x}_i)] + (1 - y_i) \ln [1 - \pi(\mathbf{x}_i)]\} \quad (16)$$

2.2.2. Uji Signifikansi Model

Pengujian signifikansi parameter model regresi logistik dilakukan dengan dua cara yaitu pengujian secara simultan dan parsial. Pengujian secara simultan dengan hipotesis sebagai berikut :

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_n = 0$$

$$H_1: \text{paling sedikit ada satu } \beta_i \neq 0$$

Statistik uji G :

$$-2 \ln \left[\frac{L_0}{L_1} \right] \sim \chi_n^2 \quad (17)$$

L_0 = *likelihood* tanpa peubah bebas

L_1 = *likelihood* dengan peubah bebas

n = derajat bebas

Statistik uji G mengikuti sebaran χ_n^2 . H_0 ditolak jika nilai statistik uji $G > \chi_n^2$.

Pengujian secara parsial dengan hipotesis sebagai berikut :

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Statistik uji *Wald* :

$$\frac{\hat{\beta}_i}{Se(\hat{\beta}_i)} \sim Z \quad (18)$$

H_0 ditolak jika $|W_i| > Z$ artinya peubah penjelas berpengaruh nyata terhadap peubah respon, dengan Z mengikuti sebaran normal baku (Hosmer dan Lemeshow 2000).

2.3. Optimalisasi Ketepatan Klasifikasi

2.3.1. Confusion matrix

Confusion matrix adalah sebuah tabel yang terdiri dari banyaknya baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi. Tabel ini diperlukan untuk mengukur kinerja suatu model klasifikasi.

Tabel 1 *Confusion matrix*

		Prediksi	
		0	1
Aktual	0	<i>TN</i>	<i>FP</i>
	1	<i>FN</i>	<i>TP</i>

$$\text{Sensitifitas (\%)} = \frac{TP}{TP+FN}$$

$$\text{Spesifisitas (\%)} = \frac{TN}{TN+FP}$$

$$\text{Akurasi (\%)} = \frac{TN+TP}{TN+FP+FN+TP}$$

Keterangan :

TN = negatif sebenarnya (*true negative*)

FP = positif palsu (*false positive*)

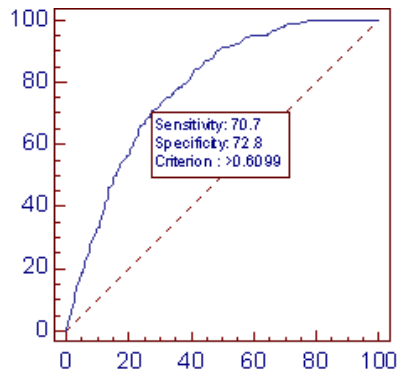
FN = negatif palsu (*false negative*)

TP = positif sebenarnya (*true positive*)

2.3.2 Receiver Operating Characteristic (ROC)

ROC adalah grafik yang menggambarkan kinerja sistem klasifikasi biner antara sensitifitas pada sumbu Y dengan 1-spesifisitas pada sumbu X . Sensitifitas adalah ukuran

ketepatan klasifikasi dari suatu kejadian yang diharapkan sedangkan spesifisitas adalah ukuran ketepatan klasifikasi dari suatu kejadian yang tidak diharapkan. Tujuan ROC untuk menentukan *cut off point* pada uji diagnostik yang bersifat kontinu.



Gambar 6 *Receiver Operating Characteristic (ROC)*

Gambar 6 menunjukkan kurva ROC dengan tingkat sensitifitas 70.7%, spesifisitas 72.8% dan nilai *cut off point* sebesar 0.6099. Penentuan *cut off point* ditandai dengan tingkat sensitifitas dan spesifisitas yang tinggi pada kurva ROC.

Indikasi keseluruhan akurasi diagnostik kurva ROC menggunakan nilai *Area Under Curve* (AUC). Nilai AUC dapat dibagi menjadi beberapa kelompok (Gorunescu, 2011) sebagai berikut :

Tabel 2 Klasifikasi data berdasarkan nilai AUC

AUC	Klasifikasi
0.90-1.00	Sangat baik
0.80-0.90	Baik
0.70-0.80	Cukup baik
0.60-0.70	Kurang baik
0.50-0.60	Tidak baik

3. HASIL DAN PEMBAHASAN

3.1. Data Penelitian

Data penelitian ini merupakan pengajuan aplikasi nasabah kredit tanpa agunan (KTA) pada Bank Internasional Indonesia (BII) bulan Mei 2014 yang terdiri dari 1700 pengamatan, 18 peubah *input* dan 1 peubah target *output*. *Software* yang digunakan penelitian ini adalah *Matlab* 7.0.1, *R2.14.0*, *Minitab* 14.

Tabel 3 Data pengajuan aplikasi nasabah kredit KTA

Keterangan	Skala pengukuran	Kategori	Kode						Peubah
			1	2	3	4	5	6	

Umur	Rasio								X_1
Jenis kelamin	Nominal	Pria	1						X_2
		Wanita	0						
Status pernikahan	Nominal	Belum menikah	1	0					X_3
		Menikah	0	1					X_4
		Janda/duda	0	0					
Jumlah tanggungan	Rasio								X_5
Pendidikan	Ordinal	SD	1	0	0	0	0	0	X_6
		SMP	0	1	0	0	0	0	X_7
		SMA	0	0	1	0	0	0	X_8
		Akademi	0	0	0	1	0	0	X_9
		S1	0	0	0	0	1	0	X_{10}
		S2	0	0	0	0	0	1	X_{11}
		S3	0	0	0	0	0	0	
Pekerjaan	Nominal	Karyawan swasta	1	0	0				X_{12}
		Pegawai negeri	0	1	0				X_{13}
		Wiraswasta	0	0	1				X_{14}
		Profesional	0	0	0				
Riwayat nasabah	Nominal	Dapat mengajukan	1						X_{15}
		Tidak dapat mengajukan	0						
Jumlah cicilan	Ordinal	12 bulan	1	0					X_{16}
		24 bulan	0	1					X_{17}
		36 bulan	0	0					
Standar gaji	Nominal	Sesuai standar minimum	1						X_{18}
		Tidak sesuai standar minimum	0						
Keputusan	Nominal	Approve	1						Y
		Reject	0						

3.2. Metode Analisis

Data kredit dipisahkan menjadi dua bagian yaitu data *training* dan data *testing*. Pembagian data ini dilakukan secara acak. Data *training* sebesar 70% untuk mendapatkan model dan data *testing* sebesar 30% untuk proses validasi.

Langkah-langkah yang dilakukan untuk klasifikasi ANN *Backpropagation* adalah sebagai berikut :

1. Penentuan jumlah lapisan

Lapisan *input*, lapisan tersembunyi, dan lapisan *output* pada data *training*.

2. Penentuan bobot.

Lapisan *input* ke lapisan tersembunyi, dan lapisan tersembunyi ke lapisan *output*.

3. Penentuan fungsi aktivasi

Fungsi aktivasi dari lapisan *input* ke lapisan tersembunyi dan lapisan tersembunyi ke lapisan *output*.

4. Penentuan nilai *learning rate*

5. Proses pembelajaran ANN *Backpropagation*

Algoritma *Backpropagation* untuk menghasilkan bobot akhir. Hasil bobot akhir akan digunakan pada proses validasi model.

6. Penentuan optimasi ketepatan klasifikasi

Confusion matrix dan kurva ROC.

Langkah-langkah yang harus dilakukan untuk klasifikasi regresi logistik adalah sebagai berikut :

1. Uji signifikansi koefisien regresi

Statistik uji *G* untuk uji signifikansi koefisien regresi secara simultan dan secara parsial dengan statistik uji *wald*.

2. Penentuan optimasi ketepatan klasifikasi

Confusion matrix dan kurva ROC.

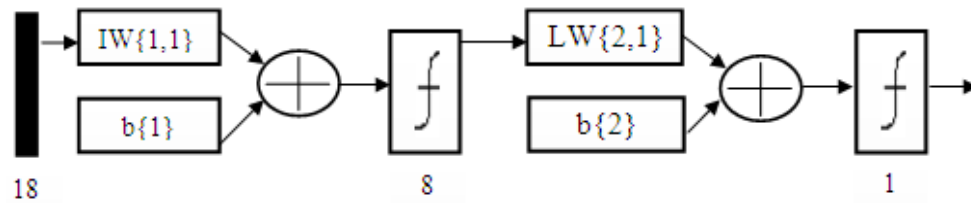
3.3. Pembahasan

3.3.1. Pendugaan Model ANN *Backpropagation*

Pembentukan model ANN *Backpropagation* pada data *training* dengan jumlah unit lapisan *input* adalah 18 unit, lapisan unit *output* 1 unit sedangkan penentuan lapisan unit tersembunyi, fungsi aktivasi, dan *learning rate* (α) akan dilakukan simulasi berdasarkan nilai KTG yang minimum. Hasil simulasi menghasilkan 8 unit lapisan tersembunyi, fungsi aktivasi *tangen sigmoid* pada lapisan *input* ke tersembunyi dan fungsi aktivasi linear pada lapisan tersembunyi ke *output*, α sebesar 0.8. Jadi model ANN *Backpropagation* yang didapatkan adalah sebagai berikut :

$$y_k = f_k \left(\sum_{j=1}^8 w_{jk} f_j \left(v_{0j} + \sum_{i=1}^{18} x_i v_{ij} \right) + w_{0k} \right)$$

Jumlah 18 unit *input* terdiri dari umur, jenis kelamin, status menikah, janda/duda, jumlah tanggungan, pendidikan SD, SMP, SMA, Akademi, S1, S2, karyawan swasta, pegawai negeri, wiraswasta, riwayat nasabah, jumlah cicilan 12 dan 24 bulan, standar gaji.



Gambar 8 Hasil simulasi Artificial Neural Network *Backpropagation*

Gambar 8 menunjukkan bahwa hasil simulasi yang telah didapatkan dengan 18 unit lapisan *input*, 8 unit lapisan tersembunyi, 1 unit lapisan *output*, $IW\{1,1\}$ dan $b\{1\}$ adalah bobot, bias akhir yang terdapat pada lapisan *input* ke lapisan tersembunyi, $LW\{2,1\}$ dan $b\{2\}$ adalah bobot, bias akhir yang terdapat pada lapisan tersembunyi ke lapisan *output*.

Tabel 4 Bobot dan bias lapisan *input* ke tersembunyi

Bias (v_{0j})	Bobot (v_{ij})				
	1	2	3	4	5
64.3094	-3.6348	-12.8357	-131.1289	-72.2237	-57.3722
91.9805	-1.8816	21.983	-6.6218	7.6383	0.2872
-75.4847	0.0161	-0.4806	4.0478	4.8925	-0.2791
111.5382	-2.3915	19.5151	23.9073	73.9272	-4.0328
23.7452	-0.6509	-3.3627	39.9551	11.0554	1.4695
20.7293	-0.0108	-3.6996	-5.7927	-9.4565	1.3732
-52.1367	30.4415	-120.5041	49.5338	-216.5526	-217.884
35.9531	0.1779	3.7001	-24.4186	-26.9846	0.1363

Tabel 4 merupakan bobot v_{ij} dan bias v_{0j} akhir pada lapisan *input* dengan 5 unit yang terdiri dari umur, jenis kelamin, status menikah, janda/duda, jumlah tanggungan dan lapisan tersembunyi dengan 8 unit, untuk nilai bobot yang lebih lengkap terdapat pada lampiran 3.

Tabel 5 Bobot dan bias lapisan tersembunyi ke *output*

Bobot dan bias	Nilai
w_{jk}	-0.0405
	-0.1253
	0.2381
	0.0392
	-0.0543
	-0.1642
	0.0384

	0.2503
w_{0k}	0.4238

Tabel 5 merupakan nilai bobot w_{jk} dan bias w_{0k} akhir pada lapisan tersembunyi ke lapisan *output* yang didapatkan dari hasil simulasi. Nilai bobot dan bias yang didapatkan pada proses data *training* akan digunakan untuk validasi data *testing*.

3.3.2. Pendugaan Model Regresi Logistik

Model akhir pada regresi logistik adalah sebagai berikut:

$$\ln\left(\frac{\hat{\pi}(\mathbf{x})}{1 - \hat{\pi}(\mathbf{x})}\right) = -1.4167 + 0.4799 x_2 + 1.1512 x_{16} + 0.9564 x_{17} + 0.8139 x_{18}$$

$$\hat{\pi}(\mathbf{x}) = \frac{\exp(-1.4167 + 0.4799 x_2 + 1.1512 x_{16} + 0.9564 x_{17} + 0.8139 x_{18})}{1 + \exp(-1.4167 + 0.4799 x_2 + 1.1512 x_{16} + 0.9564 x_{17} + 0.8139 x_{18})}$$

Berdasarkan model akhir di atas, dari 18 peubah penjelas pada tabel 3, hanya terdapat 4 peubah penjelas yang signifikan pada model regresi logistik yaitu jenis kelamin, jumlah cicilan 12 dan 24 bulan, standar gaji. Jadi model ini yang akan digunakan untuk proses validasi data *testing* di regresi logistik.

Uji signifikansi parameter model regresi logistik dapat dilihat pada Tabel 6 sebagai berikut :

Tabel 6 Uji signifikansi dan *odds ratio* model regresi logistik

Peubah penjelas	Koefisien	Z	P-value	G	P-value	Odds ratio
Konstanta	-1.4167	-3.83	0.00	83.064	0.00	
Jenis kelamin	0.4799	3.66	0.00			1.62
Jumlah cicilan 12 bulan	1.1512	5.83	0.00			3.16
Jumlah cicilan 24 bulan	0.9564	7.25	0.00			2.60
Standar gaji	0.8139	2.32	0.02			2.26

Tabel 6 menunjukkan bahwa peubah penjelas yang memiliki pengaruh signifikan terhadap peubah respon hasil keputusan secara parsial yaitu jenis kelamin, jumlah cicilan 12 dan 24 bulan, standar gaji. Secara simultan peubah penjelas jenis kelamin, jumlah cicilan 12 dan 24 bulan serta standar gaji dapat menjelaskan peubah respon hasil keputusan.

Interpretasi model regresi logistik dapat dijelaskan dengan *odds ratio* sebagai berikut :

1. $Odds\ ratio_1 = Odds$ nasabah *approve* pada pria 1.62 kali dari *odds* nasabah *approve* pada wanita
2. $Odds\ ratio_2 = Odds$ nasabah *approve* yang memiliki jumlah cicilan 12 bulan 3.16 kali dari *odds* nasabah *approve* yang memiliki jumlah cicilan 36 bulan.
3. $Odds\ ratio_3 = Odds$ nasabah *approve* yang memiliki jumlah cicilan 24 bulan 2.60 kali dari *odds* nasabah *approve* yang memiliki jumlah cicilan 36 bulan.
4. $Odds\ ratio_4 = Odds$ nasabah *approve* yang memiliki gaji sesuai standar minimum 2.26 kali dari *odds* nasabah *approve* yang memiliki gaji tidak sesuai standar minimum.

3.3.3. Optimasi Ketepatan Klasifikasi

Optimasi ketepatan klasifikasi berdasarkan *confusion matrix* dan ROC pada model ANN *Backpropagation* dan regresi logistik adalah sebagai berikut.

Tabel 8 *Confusion matrix* data training

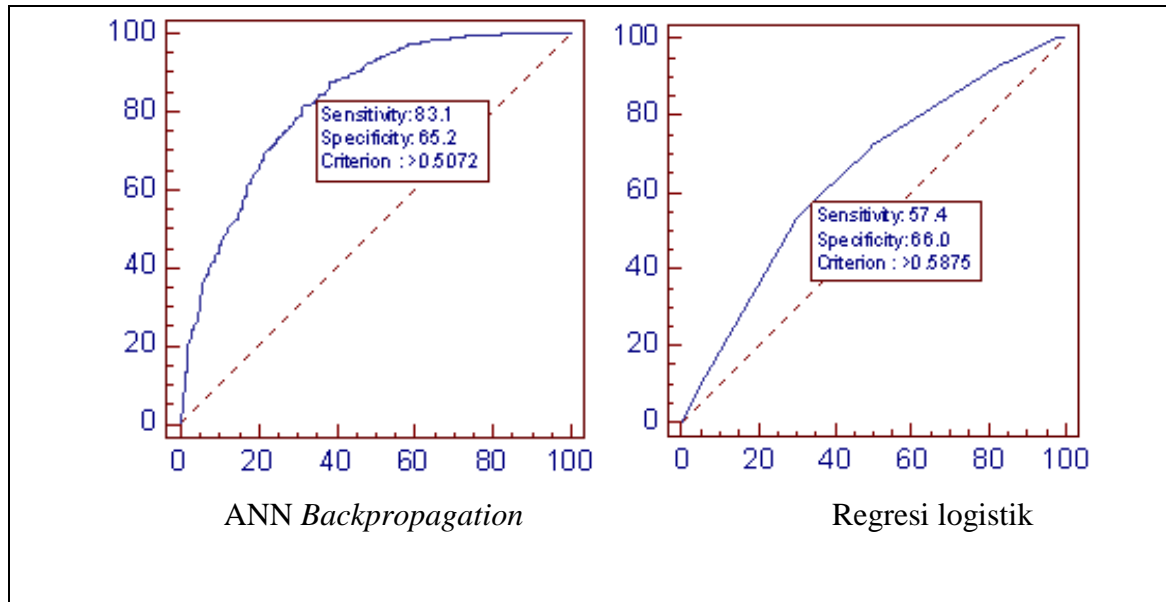
Aktual	ANN		Regresi logistik	
	<i>Backpropagation</i>			
	<i>Reject</i>	<i>Approve</i>	<i>Reject</i>	<i>Approve</i>
<i>Reject</i>	65.19	34.81	66.00	34.00
<i>Approve</i>	16.88	83.12	42.57	57.43

Tabel 8 menunjukkan bahwa prediksi model ANN *Backpropagation* dengan *approve* 83.12% dan *reject* 65.19%, regresi logistik dengan hasil *approve* 57.43% dan *reject* 66.00%.

Tabel 9 *Confusion matrix* data testing

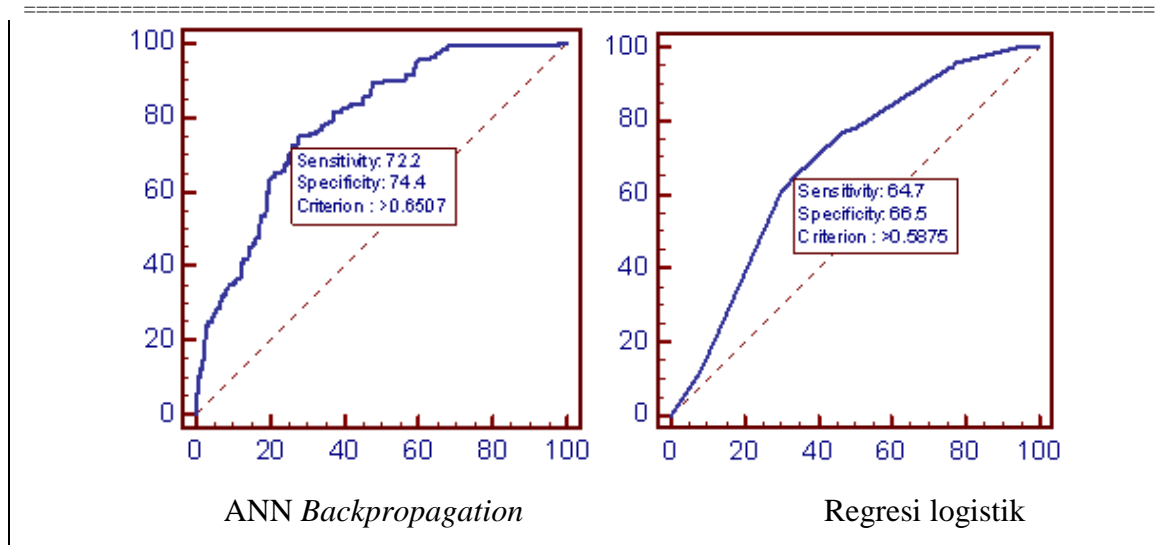
Aktual	ANN		Regresi logistik	
	<i>Backpropagation</i>			
	<i>Reject</i>	<i>Approve</i>	<i>Reject</i>	<i>Approve</i>
<i>Reject</i>	73.95	26.05	66.51	33.49
<i>Approve</i>	27.80	72.20	35.25	64.75

Tabel 9 menunjukkan bahwa prediksi model ANN *Backpropagation* dengan *approve* 72.20% dan *reject* 73.95%, regresi logistik dengan hasil *approve* 64.75% dan *reject* 66.51%.



Gambar 9 ROC data *training* pada model ANN *Backpropagation*, regresi logistik

Gambar 9 menunjukkan bahwa *cut off point* yang dihasilkan pada model ANN *Backpropagation* dan regresi logistik pada data *training* masing-masing sebesar 0.5072 dan 0.5875. *Area under curve* (AUC) pada model ANN *Backpropagation* sebesar 0.8250, regresi logistik 0.6450. Berdasarkan nilai AUC dari ke dua model tersebut model ANN *Backpropagation* adalah model yang terbaik, sehingga dapat disimpulkan bahwa akurasi diagnostik klasifikasi *approve* dan *reject* pada nasabah KTA dapat diklasifikasikan dengan baik sebesar 82.50% bila menggunakan model ANN *Backpropagation*.



Gambar 10 ROC data *testing* pada model ANN *Backpropagation*,
regresi logistik

Gambar 10 menunjukkan bahwa *cut off point* yang dihasilkan model ANN *Backpropagation* dan regresi logistik pada data *testing* yaitu 0.6507, 0.5875. Area under curve (AUC) pada model ANN *Backpropagation* sebesar 0.7910, regresi logistik 0.6850, dari kedua model tersebut nilai AUC ANN *Backpropagation* adalah yang terbaik, sehingga dapat disimpulkan bahwa akurasi diagnostik klasifikasi *approve* dan *reject* pada nasabah KTA dapat diklasifikasikan secara cukup baik sebesar 79.10% dengan menggunakan model ANN *Backpropagation*.

Nilai AUC pada data *training* dan *testing* model ANN *Backpropagation* adalah model yang terbaik pada pengklasifikasian nasabah KTA diikuti oleh model regresi logistik. Jadi dari hasil yang didapatkan maka model ANN *Backpropagation* adalah model yang terbaik dibandingkan regresi logistik pada pengklasifikasian nasabah KTA. Berdasarkan tingkat sensitifitas dan spesifisitas pada data *training* dan data *testing*, metode ANN *Backpropagation* memiliki tingkat sensitifitas dan spesifisitas lebih baik diikuti oleh regresi logistik.

Tabel 10 Akurasi Model

Model	Akurasi (%)	
	Data <i>training</i>	Data <i>testing</i>
ANN <i>Backpropagation</i>	75.63	72.94
Regresi logistik	61.01	65.49

Tabel 10 model yang memiliki nilai akurasi terbaik pada data *training* dan data *testing* yaitu model ANN *Backpropagation* diikuti model regresi logistik.

Berdasarkan tingkat ketepatan klasifikasi pada *confusion matrix* data *training* dan *testing* model ANN *Backpropagation* memiliki ketepatan klasifikasi nasabah *approve* dan *reject* yang cukup baik diikuti oleh regresi logistik. Berdasarkan keakurasian model, ANN

Backpropagation memiliki akurasi terbaik pada data *training* dan data *testing* diikuti oleh model regresi logistik.

Penentuan peubah penjelas yang berpengaruh terhadap model cukup sulit bila diterapkan dengan metode ANN *Backpropagation*, dikarenakan ANN *Backpropagation* dibangun oleh bobot-bobot yang cukup banyak dalam pembentukan model, walaupun demikian ANN *Backpropagation* memiliki kemampuan yang cukup baik dengan diikuti sertakan seluruh peubah penjelas dalam pembentukan model.

Proses pembentukan model regresi logistik lebih cepat dibandingkan ANN *Backpropagation*, dikarenakan proses simulasi yang cukup lama untuk mendapatkan bobot-bobot yang optimum, akan tetapi ANN *Backpropagation* memiliki akurasi klasifikasi yang cukup baik dibandingkan dengan regresi logistik.

4. KESIMPULAN

Metode regresi logistik memiliki kemampuan untuk menentukan peubah penjelas yang berpengaruh terhadap peubah respon hasil keputusan. Regresi logistik dengan peubah penjelas berpengaruh yaitu jenis kelamin, jumlah cicilan 12 bulan, jumlah cicilan 24 bulan, dan standar gaji sedangkan. Jadi pihak bank dapat menjadikan peubah penjelas tersebut sebagai pertimbangan untuk menentukan hasil keputusan nasabah KTA. Berdasarkan nilai ketepatan klasifikasi pada *confusion matrix*, nilai akurasi, dan AUC metode yang terbaik pada data nasabah KTA yaitu ANN *Backpropagation* diikuti oleh regresi logistik.

DAFTAR PUSTAKA

- Bank Internasional Indonesia. 2014. Data Aplikasi Nasabah Kredit Tanpa Agunan Mei 2014. Jakarta : Bank Internasional Indonesia.
- Cox DR, Snell EJ. 1989. *Analysis of Binary Data Second Edition*. London: Chapman & Hall.
- Gorunescu, F. 2011. *Data Mining : Concepts, Models, and Techniques*. Verlag Berlin Heidelberg : Berlin
- Hosmer DW, Lemeshow S. 2000. *Applied Logistic Regression*. John Wiley and Sons : New York.
- Haykin, S. 1999. *Neural Networks : A Comprehensive Foundation (2nd ed)*. Upper Saddle River, NJ : Prentice-Hall, Inc.
- Mirtalaei MS, Saberi M, Hussain OK, Ashjari B, Hussain FK . 2012. A Trust-based Bio-inspired Approach for Credit Lending Decisions. *Journal of Computing*. 94:541-577.
- Reyson PR, Donna MSR. 2012. Artificial Neural Network Model and Multiple Regression Analysis Model in Predicting Rainfall-The Case of Isabela, Philippines. *Journal of Arts and Sciences*. 5(4) : 243-263.
- Shi HY, Lee KT, Lee HH, Ho WH, Sun DP, Wang JJ, Chiu CC. 2012. Comparison of Artificial Neural Network and Logistic Regression Models for Predicting In-Hospital Mortality after Primary Liver Cancer Surgery. *Journal of Statistics*. 7: e35781.
- Sutikno. 2008. Statistical Downscaling Luaran GCM dan Pemanfaatannya untuk Peramalan Produksi Padi. [Disertasi]. Bogor : Sekolah Pascasarjana, Institut Pertanian Bogor.